OsteoArthritis and Cartilage (2007) **15,** 918–931

doi:10.1016/j.joca.2007.02.003

Osteoarthritis and Cartilage

International Cartilage Repair Society

OARSI OSTEOARTHRITIS RESEARCH SOCIETY INTERNATIONAL

Responsiveness and minimally important change for the Manchester-Oxford foot questionnaire (MOXFQ) compared with AOFAS and SF-36 assessments following surgery for hallux valgus

J. Dawson D.Phil., Senior Research Scientist†*[a], H. Doll D.Phil., Medical Statistician†, J. Coffey M.Sc., Research Officer‡, C. Jenkinson D.Phil., Professor of Health Services Research† and on behalf of the Oxford and Birmingham Foot and Ankle Clinical Research Group

† *Department of Public Health, University of Oxford, Old Road Campus, Headington, Oxford, UK*
‡ *School of Health and Social Care, Oxford Brookes University, Marston Road Campus, Jack Straws Lane, Oxford, UK*

Summary

Objectives: To assess responsiveness and minimally important change (MIC) for the Manchester-Oxford foot questionnaire (MOXFQ) using anchor and distribution-based approaches. Responsiveness and estimates of minimal clinically important difference (MCID) and minimal detectable change are compared with those from the Short-Form 36 (SF-36) and American Orthopaedic Foot & Ankle Society (AOFAS) measures.

Methods: A prospective observational study of 91 consecutive patients (125 foot operations) undergoing hallux valgus surgery at an orthopaedic hospital. Pre- and 12 month post-surgery, patients completed the MOXFQ and SF-36, and foot surgeons assessed all four AOFAS scores corresponding to four regions of the foot. Transition items were asked about perceived changes compared with before surgery.

Results: Mean changes in all domains of each instrument were statistically significant, but foot-specific MOXFQ and AOFAS domains produced much larger effect sizes (>1) than any SF-36 domains, indicating superior responsiveness. Clear associations occurred between transition items and all MOXFQ and AOFAS scores, but with only one (physical function) SF-36 domain. Anchor and distribution-based approaches identified generally comparable measures of MIC, which for the MOXFQ and AOFAS domains were between 1 and 2 standard error of measurement. In metric terms, the MCIDs were 16, 12, and 24 for the MOXFQ Walking/standing, Pain, and Social Interaction domains, respectively.

Conclusions: For hallux valgus surgery, the MOXFQ is highly responsive. Performance is comparable to the AOFAS and notably better than the generic SF-36. Study estimates of MIC for the MOXFQ are useful to inform sample-size calculations for future clinical trials.

Key words: Foot and ankle, Hallux valgus, Surgery, Quality of life, Outcomes assessment, Responsiveness, Minimal clinically important differences.

Introduction

Foot and ankle pain is very common, particularly in older women[1], being caused by many conditions[2], including arthritis[3]. A wide range of foot and ankle disorders are treated surgically, with foot and ankle surgery constituting over 15% of orthopaedic practice[4]. Many forms of foot and ankle surgery have not, however, been adequately evaluated on account of the lack of appropriate, standardised methods of outcomes assessment. In particular, there is a lack of patient-focused and patient-reported measures[5,6], measures devised with patients' input and completed by them and thus reflecting the patients' not the clinicians' perspective[7,8]. Bias is minimized by making assessment completely independent of the surgical team[9]. Indeed, The Royal College of Surgeons has recommended that validated patient-reported outcome measures should be used, in preference to clinical assessments, as the primary outcome in clinical trials[10].

One such patient-focused and patient-reported instrument, the Manchester-Oxford foot questionnaire (MOXFQ)[11], has been developed for use as an outcome measure for all types of foot surgery. We recently reported on the development and initial assessment of the MOXFQ in patients undergoing surgery for hallux valgus; a common condition, affecting up to a third of the population[1,12] and for which clinical trials of surgical interventions for the condition are urgently needed[5,13].

[a]Visiting Professor at Oxford Brookes University.

*Address for correspondence and reprint requests to: Dr Jill Dawson, Senior Research Scientist, Department of Public Health, University of Oxford, Old Road Campus, Headington, Oxford OX3 7LF, UK. Tel: 44-(0)1865-289423; Fax: 44-(0)1865-289436; E-mail: jill.dawson@dphpc.ox.ac.uk

Received 9 October 2006; revision accepted 4 February 2007.

In prospective outcome studies, such as a trial, the responsiveness of an outcome measure, that is, its ability accurately to detect change when it has occurred[14], is the most important aspect[15]. Nevertheless, while it is simple to determine the statistical significance of any such changes, it can be harder to determine the real clinical or subjective meaning of these changes, particularly given the partial dependence of statistical significance on sample size. This problem can be largely overcome if the minimally important change (MIC) is known for the measure of interest[16]. The MIC can act as a yardstick for clinical relevance, so that changes in scores exceeding the MIC are known to be clinically relevant[17].

There are essentially two approaches to determining the smallest amount of change on a measure that is likely to be of importance. The first approach is distribution based, i.e., based on the statistical characteristics of the sample under study. Examples include the effect size (ES), the standardised response mean (SRM), the standard error of measurement (S.E.M.), and the minimal detectable change (MDC)[16,18]. The ES and SRM represent change in standard deviation (SD) units, whereas the S.E.M. and MDC are provided in the same units as the original measurement. This method provides the smallest change which, *for an individual*, is likely to be beyond the measurement error of the instrument and thus to represent a true change. However, it is not known whether this change is clinically relevant. The second approach uses anchor-based methods, where an external criterion is used to provide an indication of the minimal change that is of clinical relevance *at the group level*. The most commonly used such measure is the minimal clinically important difference (MCID) which is defined as the smallest change in the measure that patients perceive as meaningful and which would cause clinicians to consider a change in the patient's management[19]. Another measure is based on diagnostic test methodology to give a best cut-point[20]. The advantage of this approach is that the concept of 'minimal importance' is here explicitly defined, which is not the case with the former approach, although there is no indication whether the change lies within the measurement error of the instrument[17]. The two different approaches are therefore different and unlikely to produce identical results.

The purpose of the analyses reported in this paper is twofold: (1) to examine various forms of evidence for the responsiveness of the MOXFQ; and (2) to ascertain the MIC for each of the three scales of the MOXFQ. In each case, findings for the MOXFQ are compared with equivalent results for the American Orthopaedic Foot & Ankle Society (AOFAS) scores and the generic, patient-reported, Short-Form 36 (SF-36).

Materials and methods

Local ethics committee approval was obtained (Applied and Qualitative Research Ethics Committee reference A02.009) and all subjects consented to participate in the study.

SUBJECTS

Adult patients, aged between 20 and 75 years, were consecutively recruited from an orthopaedic hospital within 4 weeks prior to surgery for hallux valgus. Details of the recruitment procedures and sample size have been reported elsewhere[11].

ASSESSMENTS

Patients completed a pre-operative questionnaire, unassisted, while attending a hospital pre-admission clinic. This contained (1) demographic items, (2) the MOXFQ[11]—completed separately for each foot to be operated on; and (3) the SF-36 general health questionnaire[21,22].

The MOXFQ[11] (see Appendix 1) contains 16 items, each with five response options, comprising three separate underlying dimensions: foot pain (five items), walking/standing problems (seven items) and issues related to social interaction (four items), including feelings of self-consciousness about foot/footwear appearance ('cosmesis'). Item responses are each scored from 0 to 4, with 4 representing the most severe state. The scale score representing each dimension is produced by summing the responses to each item within that dimension. Raw scale scores are then converted to a metric (0−100; 100 = most severe).

The SF-36 contains 36 items and is widely used as a generic health status instrument. It provides scores on eight dimensions of health: physical functioning, social functioning, role limitations due to physical problems, role limitations due to emotional problems, mental health, energy/vitality, bodily pain and general health perceptions over the last 4 weeks. Scores for each dimension are produced by summing the responses to each item within that dimension and then converting the raw scores to a metric (0−100, 100 = good health).

The MOXFQ and SF-36 assessments were repeated in an out-patient clinic setting 12 months post-surgery. Two additional 'transition' items were added on this occasion, asking about perceived changes compared with before foot surgery (see Table III for details) as well as three satisfaction items: (1) 'Overall, how pleased are you now with the result of your surgery on your foot?', (2) 'Compared with before your operation, how pleased are you with: (1) the appearance of your foot?', (2) the range of shoes that you can wear?'. Response options to all three items were: 'very pleased', 'fairly pleased', 'not very pleased', 'very disappointed'.

In addition, the foot surgeon (a member of the Oxford and Birmingham Foot and Ankle Clinical Research Group) completed pre- and 12 month post-surgical clinical examinations using all four ratings of the standard AOFAS[23]. These correspond to four regions of the foot: the ankle and hindfoot, the midfoot, the hallux metatarsophalangeal (MTP)−interphalangeal (IP) region and the lesser toe MTP−IP region. The four domain scales are produced by applying domain-specific weightings to individual item scores and then summing the weighted item scores within each domain. Thus, while some identical items appear in more than one scale, they do not necessarily have the same score weighting. Nonetheless, for each scale, the minimum score is 0 and the maximum score 100, with 100 representing the best health. A score of 100 would be obtained, for each scale, by a patient with no pain, full range of motion, no instability, good alignment, no limitation of daily or recreational activities, and no footwear limitations for that particular foot region. The assessor was blinded to patients' responses to the other health status instruments.

Additional information was obtained pre-operatively, and again at 3 months following surgery, concerning two measures of the hallux valgus deformity taken from routine radiographs of the patient's foot, using a goniometre: (1) the hallux abductus angle, measured as the angle formed between the longitudinal bisections of the first metatarsal and proximal phalanx, and (2) the first–second intermetatarsal angle, measured as the angle formed between longitudinal bisections of the first and second metatarsal shafts. Methods for taking these measurements have been illustrated elsewhere[24], together with evidence of high inter-rater reliability for the hallux abductus angle and intermetatarsal angle. In our own study just one person (G.L.) provided these measures.

STATISTICAL METHODS

Data analysis was undertaken within SPSS version 13.0[25] and STATA version 8.2[26]. Data are presented as mean (SD) at each assessment and mean change (SD) from pre-admission to 12-months post-surgery, with paired *t* tests being used to assess the statistical significance of any change. Continuous data were categorized where appropriate. Statistical significance was taken at the 5% level throughout, with 95% confidence intervals (CIs) used to express uncertainty around the estimates.

Comparative assessments of responsiveness and change for each instrument were made using ES, relative efficiency, MCID, best cut-point using diagnostic test methodology, and MDC. All these measures are described below.

Responsiveness to change was first assessed using ES and overall relative efficiency. ES represents the extent of change identified by an instrument in a unitless way so allowing direct comparison between instruments[27]. The ES is calculated by dividing the mean change in scores (post-surgical minus pre-surgical scores), by the SD at time one (pre-surgical score). An ES of 1.0 is thus equivalent to a change of one SD in the sample. Values of 0.2, 0.5 and 0.8 are typically regarded as indicating small, medium and large degrees of change[27,28].

The overall relative efficiency (how efficient one measure is at capturing change relative to another measure) of the MOXFQ vs (1) the AOFAS clinical assessment and (2) the SF-36, was calculated by taking the square of the ratio of the mean (over all scales within an instrument) of the individual paired *t* statistics comparing the baseline and 12 month follow-up scores[29].

Non-parametric correlations (Spearman's rho, r_s)[30] were used to examine associations between change scores of clinical (radiographically derived hallux abductus and intermetatarsal angles and clinically assessed AOFAS scales), as well as the patient-reported transition and satisfaction items. A correlation coefficient of 0.4–0.59 was taken to indicate a moderate correlation and 0.6–0.79 a strong correlation.

Relationships between changes in the MOXFQ, AOFAS and SF-36 scales between pre- and 12 months post-surgery were further examined by comparing mean score changes and ES in relation to different levels of response to the patient-reported transition items at 12 months following surgery. Patients who reported greater amounts of change would be expected to have correspondingly greater changes in outcome scores and associated larger ES if the outcome measures were responsive. Analysis of variance with tests for linear trend were used to assess the statistical significance and directionality of any association.

DISTRIBUTION-BASED MEASURES OF MIC

The distribution-based measures which were used to determine important change were the S.E.M. and the MDC. The S.E.M. is the error estimate for a single use of the questionnaire which is directly related to the reliability of the scale, as indicated by Cronbach's alpha (S.E.M. = SD × √1 − Cronbach's alpha). Cronbach's Alpha was calculated using the unweighted item scores of the MOXFQ, AOFAS and SF-36 scales. Note that while total scale scores for each instrument range from 0 to 100, the AOFAS scales are produced by summation of individually weighted items, the MOXFQ and SF-36 scales are produced by transformation of the sum of individual unweighted items.

The MDC is defined as the smallest amount of change between two time points that indicates a real change in the patient's health status[31], i.e., a change above that which would be expected by chance alone. With a conventional confidence level of 90%, the MDC was calculated as $1.65 \times \sqrt{2} \times$ S.E.M.. The interpretation of MDC_{90} is that 90% of truly stable patients will demonstrate random variation of less than this magnitude when assessed on multiple occasions[32].

ANCHOR-BASED MEASURES OF MIC

The MCID was determined for each of the MOXFQ, AOFAS and SF-36 scales as the mean change scores associated with a global rating of change (transition item) of 'slightly better'[16]. The MCID:S.E.M. ratio was calculated for each scale and averaged over each instrument to give an estimate of the relationship between these two anchor and distribution-based approaches and to compare its magnitude with that calculated in previous studies[17].

In addition, an approach based on diagnostic test methodology was also used[20]. Receiver operator characteristic (ROC) curves were used to assess each of the MOXFQ, AOFAS and SF-36 scales' efficiency in differentiating between those who did and did not experience important clinical change ('Slightly better'/'much better'/'no problems now' vs 'no change'/'worse', respectively). ROC curves plot sensitivity (*y*-axis) against 1-specificity (*x*-axis) for all possible cut-off points of the instrument. Sensitivity is defined as the number of patients that have improved ('Slightly better'/'much better'/'no problems now') divided by the number of all patients that have change scores above a cut-off point. Specificity refers to the number of patients who have not improved ('no change'/'worse') divided by the number of all patients who have change scores below this cut-off point. The most efficient cut-off value, with regard to specificity and sensitivity, is associated with the point closest to the top left hand corner of the ROC curve. Since the best cut-point 'cuts' or differentiates at the boundary between improvement and no change, the best cut-point would be expected to be lower than the value (the MCID) reported by all subjects in the 'slightly better' group. In addition, the greater the area under the ROC curve, the greater the ability of the scale to differentiate between those who did and did not have clinically important change. If the area under the curve is 0.5, this represents a test that is not predictive; the closer the area is to 1.0, the better the differentiation of the scale[31,33,34].

Results

STUDY POPULATION CHARACTERISTICS

The baseline study sample comprised 100 patients, representing 138 feet (i.e. each foot = one 'case'; 38 patients were booked to receive bilateral surgery on the same date). At baseline, the mean age of the patients was 50.03 (SD 12.9), median 52 years. Almost all (95%) were female. Nine people (13 foot operations) had their surgery cancelled or postponed, everyone else (91 patients, 125 cases) was followed up 12 months post-surgery.

PROCEDURES

The majority of procedures (107, 85.6%) involved a scarf osteotomy, while 13 (10.4%) were a chevron, and 5 (4.0%) other kinds of osteotomy. Eleven surgeons were involved.

RADIOGRAPHIC DATA

The radiographic data, pre- and post-surgery, are shown in Table I. Following surgery, there was a highly significant reduction in both the hallux abductus and intermetatarsal angles (both $P < 0.001$).

RESPONSIVENESS OF THE MOXFQ

The mean pre- and 12 month post-operative scores (with 95% CIs) for the MOXFQ, AOFAS and SF-36 domains are shown in Fig. 1, together with details of change scores (12 month score minus baseline score), paired *t* test comparisons and ES.

While mean changes on all domains were statistically significant ($P < 0.001$ for each MOXFQ and AOFAS domain; $P < 0.05$ for each SF-36 domain) each of the foot-specific domains produced much larger ES (all >1) than any of the generic SF-36 domains, indicating superior responsiveness. In terms of ES, the hallux scale of the AOFAS foot assessment appeared the most responsive of the four clinical scales with an ES of 2.46. The average *t* statistics over all domains within each instrument were 13.73, 12.40, and 4.10 for the MOXFQ, AOFAS and SF-36, respectively, giving a relative efficiency of the MOXFQ vs the AOFAS of 1.23 and vs the SF-36 of 11.21. Thus, while the MOXFQ was only slightly more efficient than the AOFAS, it was much more efficient than the SF-36.

The correlations (r_s) between radiographic measures (hallux abductus and intermetatarsal angles), MOXFQ, SF-36, the clinical assessment AOFAS domains and the transition and satisfaction items are shown in Table II.

Correlation coefficients that were at least moderate ($r_s > 4.0$) have been highlighted. The key finding here is that changes in measures of hallux valgus severity taken from radiographs did not correlate with any other measure. In addition, changes on the foot-specific (MOXFQ and AOFAS) measures were generally much more highly correlated ($r_s \geq 0.5$) with each other than with changes in the SF-36 domains, with the exception of the SF-36 physical function scale, with which they were generally moderately correlated ($0.4 < r_s < 0.5$). In terms of the transition items, the MOXFQ pain scale and the AOFAS hallux and AHF scales were all moderately, but highly significantly, correlated ($r_s = 0.5$, $P < 0.001$) with the transition item concerning change in pain. In addition, the MOXFQ pain scale was moderately correlated ($r_s = 0.48$) with the transition item concerning change in foot-related problems, and the MOXFQ social interaction scale was moderately correlated ($r_s = 0.44$) with the satisfaction item concerning the range of shoes that could be worn post-surgery. No SF-36 domain had any more than low correlation with changes in the transition and satisfaction items.

The relationships between the changes in MOXFQ, AOFAS and SF-36 domain scores and the patient-reported transition item concerning perceived change in foot-related problems at 12 months following surgery, in terms of mean changes and ES, are shown in Table III.

For the foot-specific measures (MOXFQ and AOFAS), there were clear associations between domain scores and global judgement of change, with statistically significant linear trends for increasing degree of improvement in domain score with greater degree of perceived improvement. By contrast, changes in the generic SF-36 domains, with the exception of scores on the physical function domain, showed almost no association with levels of response on this transition item.

ANCHOR-BASED MEASURES OF MIC

Minimal clinically important difference

The MCID estimates for the three MOXFQ domains of Walking/standing, Pain, and Social Interaction were 12.8, 4.6, and 20.3, respectively, with ES of 0.4, 0.2, and 0.8 (Table III). The ES were similar for the AOFAS scales with the exception of the Hallux scale which had an MCID with a notably larger ES of 2.0. While the MCIDs can be estimated for the SF-36 domains, because there is no trend in mean change scores across the transition item responses these estimates cannot be regarded as reliable.

Best cut-point using diagnostic test methodology

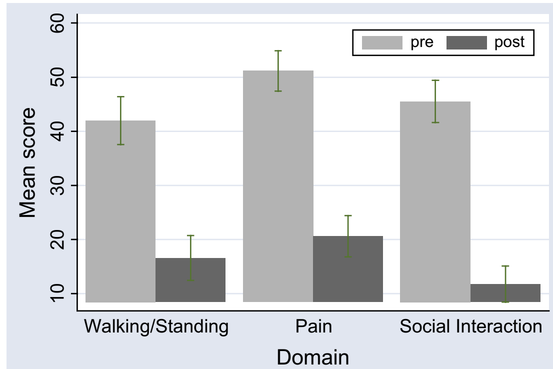
Examples of the ROC curves for the MOXFQ domains are shown in Fig. 2. These are based upon patients' responses ('Slightly better/much better/no problems' vs 'no change/worse') to the transition item concerning foot-related problems at 12 months following foot surgery.

Table IV summarises the results of all ROC curves, showing the area under the curve (with 95% CIs) and

Table I
Pre- and post-surgical hallux abductus and intermetatarsal angles, measured from radiographs. Data are shown as mean (SD) degrees

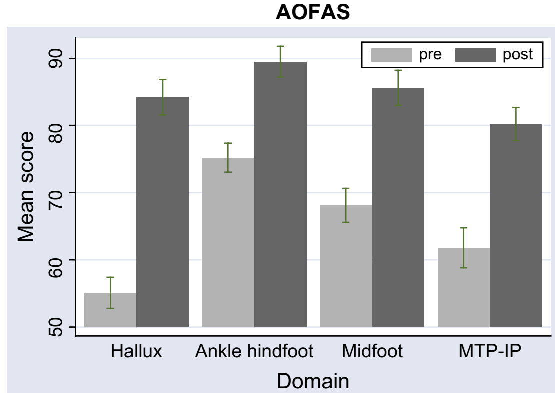
Angle	Pre-surgery	Post-surgery	Change	*P*-value
Hallux abductus	35.44 (8.96) range 18 to 60	17.93 (8.52) range 0 to 40	−17.51 (10.24) Range −3.00 to 45.00 95% CI −15.6, −19.4	<0.001
Intermetatarsal	15.26 (3.45) range 2 to 30	9.03 (2.76) range 0 to 18	−6.24 (3.41) Range −8.00 to 17.00 95% CI −5.61, −6.86	<0.001

MOXFQ



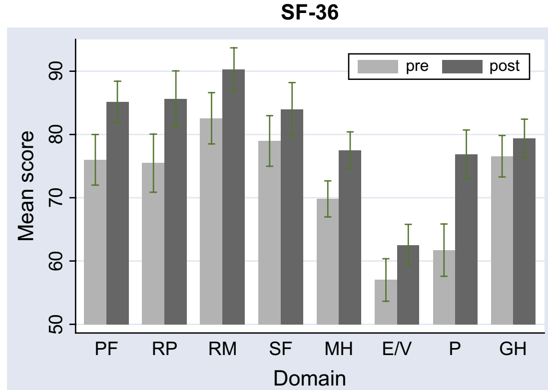
	Walking/standing	Pain	Social interaction
Pre, mean (sd)	44.9 (25.4)	52.7 (20.1)	46.6 (23.0)
Post, mean (sd)	16.4 (23.1)	19.9 (20.9)	11.8 (18.7)
Change, mean (sd)	-28.5 (27.6)	-32.8 (25.1)	-34.8 (24.1)
Paired t (p-value)	11.3 (<0.001)	14.1 (<0.001)	15.8 (<0.001)
Effect size	1.12	1.63	1.51

AOFAS



	Halux MTP-IP	Ankle hindfoot	Midfoot	Lesser toe MTP-IP
Pre, mean (sd)	54.8 (11.9)	75.3 (10.4)	68.1 (12.2)	61.4 (14.5)
Post, mean (sd)	84.2 (15.0)	90.0 (11.1)	86.1 (12.6)	79.8 (18.4)
Change, mean (sd)	29.4 (16.8)	14.7 (14.0)	18.0 (15.8)	18.4 (18.0)
Paired t (p-value)	18.1(<0.001)	10.4(<0.001)	11.3(<0.001)	9.9(<0.001)
Effect size	2.46	1.42	1.48	1.27

SF-36



	Physical function	Role Physical	Role mental	Social functioning	Mental health	Energy/ Vitality	Pain	General health
Pre, mean (sd)	75.0 (22.6)	75.1 (27.0)	83.0 (23.2)	78.3 (23.3)	70.6 (17.0)	57.3 (20.4)	61.6 (24.4)	76.2 (19.4)
Post, mean (sd)	85.3 (18.7)	85.5 (24.7)	91.0 (18.4)	84.9 (21.3)	77.5 (16.3)	62.5 (18.1)	77.2 (21.2)	79.6 (17.2)
Change,mean (sd)	10.3 (19.2)	10.4 (31.0)	7.99 (25.5)	6.6 (25.4)	6.9 (14.4)	5.2 (17.9)	15.6 (26.6)	3.5 (14.9)
Paired t (p-value)	5.6 (<0.001)	3.7 (<0.001)	3.4 (0.001)	2.8 (0.006)	5.3 (<0.001)	3.2 (0.002)	6.4 (<0.001)	2.5 (0.015)
Effect size	0.45	0.39	0.34	0.28	0.41	0.25	0.64	0.18

Fig. 1. Mean pre- and 12 month post-operative scores, change scores, ES and paired *t* test comparisons for the patient-reported MOXFQ and clinical AOFAS foot-specific assessment scales together with those from the generic SF-36 scale.

Table II

Correlation (Spearman's rho) between change scores, comparing radiographic measures (hallux abductus and intermetatarsal angles), clinical assessment (AOFAS hallux, ankle-hindfoot, midfoot, and Lesser MTP-IP scales), SF-36 (physical function, role physical, role mental, social function, mental health, energy/vitality, pain and general health perception scales) and MOXFQ (walking/standing, pain and social interaction scales); together with correlation (Spearman's rho) of these with transition and satisfaction items

	XRAY HALLUX	XRAY INTERMT	MOXFQ WS	MOXFQ PAIN	MOXFQ SI	SF-36 PF	SF-36 RP	SF-36 RM	SF-36 SF	SF-36 MH	SF-36 EV	SF-36 PAIN	SF-36 GHP	AOFAS HALLUX	AOFAS AHF	AOFAS MF	AOFAS MTP-IP
XRAY: HALLUX		0.37***	-0.08	-0.09	-0.14	0.04	0.08	-0.09	-0.08	-0.13	-0.11	-0.12	-0.07	-0.01	-0.05	0.03	0.13
INTERMT	0.37		-0.17	-0.03	-0.18	0.07	0.05	-0.002	0.04	0.02	-0.04	0.10	0.04	0.06	0.09	0.10	0.16
MOXFQ: WS	-0.08	-0.17			0.49***	-0.62***	-0.46***	-0.20*	-0.40***	-0.11	-0.07	-0.44***	-0.21*	-0.36***	-0.57***	-0.51***	-0.46***
PAIN	-0.09	-0.03	0.70***		0.42***	-0.47***	-0.24*	-0.15	-0.22*	-0.11	-0.05	-0.34***	-0.07	-0.68***	-0.76***	-0.71***	-0.68***
SI	-0.14	-0.18	0.49***	0.42***		-0.46***	-0.29**	-0.07	-0.19*	-0.06	0.09	-0.27**	-0.06	-0.36***	-0.43***	-0.37***	-0.29**
SF-36:																	
PF	0.04	0.07	-0.62***	-0.47***	-0.46***		0.63***	0.34***	0.55***	0.25**	0.19	0.53***	0.40***	0.39***	0.47***	0.43***	0.40***
RP	0.08	0.05	-0.46***	-0.24*	-0.29**	0.63***		0.45***	0.68***	0.22*	0.29**	0.55***	0.52***	0.16	0.36***	0.33**	0.21*
RM	-0.09	-0.002	-0.20*	-0.15	-0.07	0.34***	0.45***		0.56***	0.34***	0.35***	0.17	0.29**	0.08	0.22*	0.18	0.10
SF	-0.08	0.04	-0.40***	-0.22*	-0.19*	0.55***	0.68***	0.56***		0.38***	0.41***	0.40***	0.40***	0.09	0.21*	0.21	0.11
MH	-0.13	0.02	-0.11	-0.11	-0.06	0.25**	0.22*	0.34***	0.38***		0.37***	0.15	0.37***	0.16	0.14	0.11	0.14
EV	-0.11	-0.04	-0.07	-0.05	0.09	0.19	0.29**	0.35***	0.41***	0.37***		0.18*	0.26**	0.13	0.07	0.06	0.13
PAIN	-0.12	0.10	-0.44***	-0.34***	-0.27**	0.53***	0.55***	0.17	0.40***	0.15	0.18*		0.28**	0.25*	0.39***	0.36***	0.27**
GHP	-0.07	0.04	-0.21*	-0.07	-0.06	0.40***	0.52***	0.29**	0.40***	0.37***	0.26**	0.28**		0.11	-0.05	-0.07	0.07
AOFAS: HALLUX	-0.01	0.06	-0.36***	-0.68***	-0.36***	0.39***	0.16	0.08	0.09	0.16	0.13	0.25*	0.11		0.89***	0.84***	0.90***
AHF	-0.05	0.09	-0.57***	-0.76***	-0.43***	0.47***	0.36***	0.22*	0.21*	0.14	0.07	0.39***	-0.05	0.89***		0.97***	0.90***
MF	0.03	0.10	-0.51**	-0.71***	-0.37***	0.43***	0.33**	0.18	0.21	0.11	0.06	0.36***	-0.07	0.84***	0.97***		0.89***
MTP-IP	0.13	0.16	-0.46***	-0.68***	-0.29**	0.40***	0.21*	0.10	0.11	0.14	0.13	0.27**	0.07	0.90***	0.90***	0.89***	
Transition:																	
Pain now	0.07	0.01	-0.30**	-0.50***	-0.31**	0.09	-0.08	-0.16	-0.02	0.02	-0.12	-0.01	-0.10	0.45***	0.41***	0.39***	0.40***
Foot-related problems	0.16	0.09	-0.36***	-0.48***	-0.37***	0.21*	0.14	-0.08	0.07	0.06	0.02	0.19*	0.09	0.36***	0.34**	0.29**	0.30**
Satisfaction:																	
Pleased overall	0.11	0.09	-0.14	-0.32**	-0.17	0.15	-0.07	-0.04	0.08	0.02	-0.16	0.02	-0.09	0.18	0.13	0.11	0.15
Pleased with foot appearance	0.06	0.12	-0.17	-0.28**	-0.19	0.08	0.001	0.12	0.16	0.18*	0.001	-0.01	0.03	0.22*	0.19	0.11	0.20
Pleased with shoes	0.27**	0.12	-0.31**	-0.37***	-0.44***	0.27**	0.09	-0.10	0.15	0.05	-0.04	0.14	-0.07	0.34***	0.27*	0.21*	0.30**

Correlations that are at least moderate ($r > 4.0$) have been highlighted. $^*P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$.

Table III

Comparison of mean changes (with SDs) and ES, comparing the foot-specific MOXFQ and AOFAS, and generic SF-36 scales in relation to different levels of response to a patient-reported transition item about foot-related problems at 12 months following surgery to correct hallux valgus deformity

Transition item	MOXFQ scale change			AOFAS scale change			
How are the problems related to your foot now, compared with before surgery?	WS	Pain	SI	Hallux MTP−IP	Ankle hindfoot	Midfoot	Lesser toe MTP−IP
	Mean (SD) [ES]	Mean (SD) [ES]	Mean (SD) [ES]	Mean (SD) [ES]	Mean (SD) [ES]	Mean (SD) [ES]	Mean (SD) [ES]
No problems now	$n=40$* −37.23 (25.34) [1.40]	$n=39$ −43.46 (21.2) [2.04]	$n=43$ −42.79 (21.15) [2.02]	$n=34$ 35.12 (10.74) [3.44]	$n=31$ 18.16 (9.68) [1.91]	$n=32$ 21.38 (12.47) [1.86]	$n=28$ 23.75 (13.69) [1.74]
Much better	$n=59$ −31.96 (23.42) [1.46]	$n=58$ −36.21 (20.38) [1.89]	$n=59$ −36.33 (21.70) [1.56]	$n=51$ 30.59 (16.06) [2.47]	$n=47$ 17.28 (13.08) [1.63]	$n=47$ 20.66 (14.57) [1.71]	$n=47$ 20.21 (18.01) [1.33]
Slightly better	$n=12$ −12.80 (18.43) [0.39]	$n=11$ −4.55 (15.2) [0.20]	$n=12$ −20.31 (19.42) [0.79]	$n=12$ 22.17 (15.28) [2.00]	$n=10$ 4.70 (7.09) [0.59]	$n=11$ 8.36 (11.16) [0.85]	$n=10$ 12.80 (12.97) [1.02]
No change/worse	$n=9$ 12.30 (32.05) [0.43]	$n=9$ 1.11 (26.90) [0.07]	$n=9$ −9.03 (33.09) [0.38]	$n=7$ 8.29 (28.63) [0.51]	$n=6$ −3.17 (25.83) [0.22]	$n=6$ 1.67 (28.80) [0.10]	$n=11$ −1.86 (24.99) [0.11]
Total	$n=120$ −28.48 (27.58) [1.12]	$n=117$ −32.78 (25.09) [1.63]	$n=96$ −34.77 (24.06) [1.51]	$n=104$ 29.60 (16.81) [2.46]	$n=94$ 14.93 (13.94) [1.43]	$n=96$ 18.30 (15.65) [1.50]	$n=92$ 18.80 (17.96) [1.29]
P-value for linear trend	<0.001	<0.001	<0.001	<0.001	<0.001	0.001	0.001

Transition item	SF-36							
How are the problems related to your foot now, compared with before surgery?	Physical function	Role physical	Role mental	Social function	Mental health	Energy/vitality	Pain	General health
	Mean (SD) [ES]	Mean (SD) [ES]	Mean (SD) [ES]	Mean (SD) [ES]	Mean (SD) [ES]	Mean (SD) [ES]	Mean (SD) [ES]	Mean (SD) [ES]
No problems now	$n=36$ 10.56 (14.43) [0.55]	$n=41$ 12.80 (23.51) [0.53]	$n=40$ 5.83 (22.58) [0.28]	$n=40$ 6.25 (28.02) [0.32]	$n=41$ 8.05 (15.77) [0.49]	$n=41$ 5.64 (17.33) [0.34]	$n=41$ 23.04 (24.52) [1.00]	$n=40$ 5.32 (11.94) [0.38]
Much better	$n=56$ 15.36 (19.28) [0.65]	$n=58$ 12.07 (31.89) [0.45]	$n=59$ 8.76 (24.39) [0.40]	$n=57$ 8.77 (23.26) [0.37]	$n=60$ 7.25 (13.39) [0.42]	$n=59$ 4.45 (19.14) [0.22]	$n=59$ 12.05 (27.93) [0.51]	$n=53$ 2.83 (14.17) [0.13]
Slightly better	$n=9$ 2.78 (20.02) [0.10]	$n=12$ 8.85 (38.11) [0.30]	$n=12$ 2.78 (24.19) [0.15]	$n=12$ 1.04 (25.82) [0.04]	$n=12$ 2.08 (12.15) [0.14]	$n=12$ 3.65 (13.45) [0.14]	$n=12$ 15.74 (21.43) [0.56]	$n=12$ 6.67 (17.99) [0.32]
No change/worse	$n=8$ −18.13 (8.43) [1.20]	$n=9$ −9.03 (42.75) [0.33]	$n=9$ 26.85 (38.14) [0.76]	$n=7$ 0.0 (28.87) [0]	$n=9$ 5.56 (18.10) [0.35]	$n=9$ 9.72 (19.04) [0.61]	$n=7$ 1.59 (27.54) [0.06]	$n=7$ −8.43 (24.72) [0.37]
Total	$n=109$ 10.28 (19.21) [0.45]	$n=120$ 10.42 (31.02) [0.39]	$n=120$ 7.99 (25.48) [0.34]	$n=116$ 6.57 (25.40) [0.28]	$n=122$ 6.89 (14.41) [0.41]	$n=121$ 5.17 (17.89) [0.25]	$n=119$ 15.59 (26.57) [0.64]	$n=112$ 3.46 (14.86) [0.18]
P-value for linear trend	0.001	0.110	0.206	0.522	0.343	0.794	0.034	0.114

*The ‘*n*’ in each cell varies slightly across the rows reflecting an occasional missing patient’s response to an item within the health status scale, or where missing items occurred on either the pre- or post-operative clinical assessment (hence the particular overall scale would not have been computed for that individual).

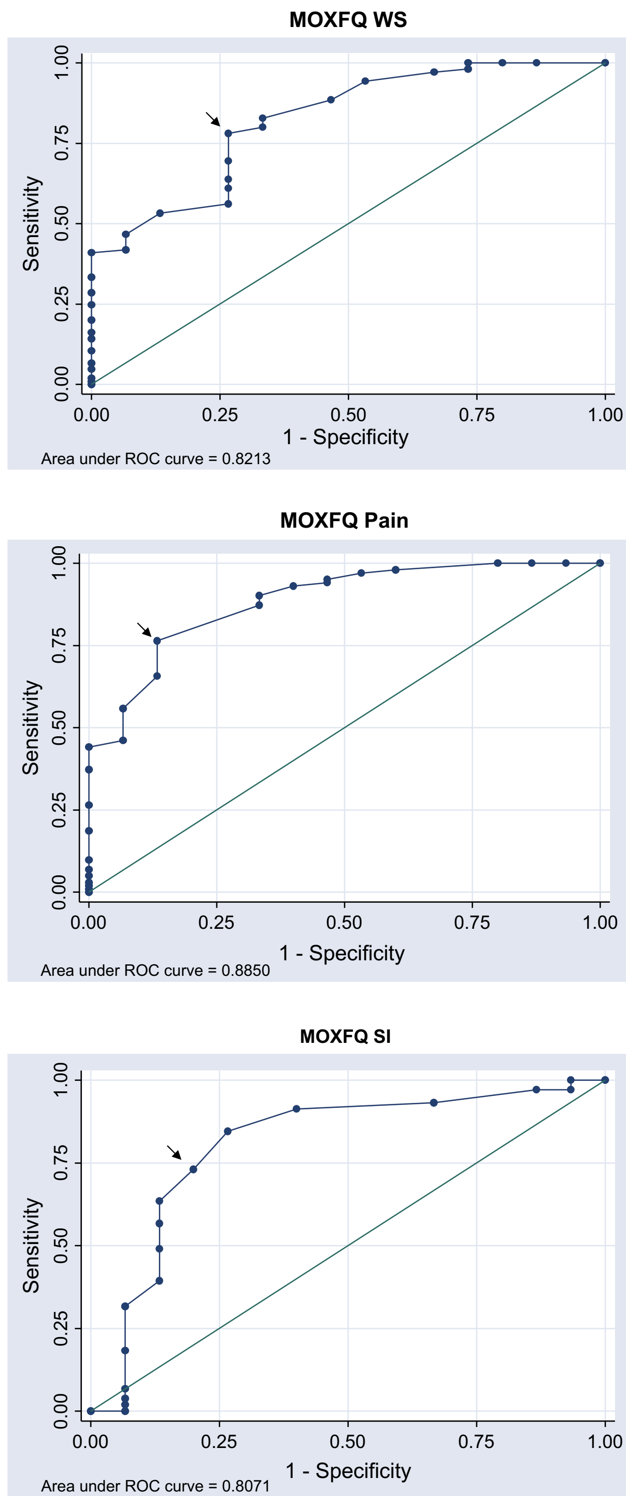


Fig. 2. ROC curves differentiating pain ‘at least slightly better’ from ‘no change/worse’ with best cut-point indicated (with an arrow) on the walking/standing, pain and social interaction scales of the MOXFQ.

best scale cut-points for both transition items. Scales having an area with 95% CIs not overlapping the value of 0.5 (therefore providing statistically significant prediction) have been highlighted. Based on the area under the curve, all three MOXFQ scales showed great ability to differentiate between patients reporting improvement vs no improvement/worse, both in terms of pain and foot-related problems. All four AOFAS scales performed well in terms of

Table IV

A comparison of ROC and best scale cut-points for the MOXFQ, AOFAS assessment and SF-36 scales, based upon responses: 'Slightly better/much better/no problems' vs 'no change/worse', to transition items about foot pain and foot-related problems, at 12 months following foot surgery

	ROC area		Best cut point	
	Pain group	**Problems**	**Pain group**	**Problems**
MOXFQ				
WS	0.82 (.71–.94)	0.87 (.73–.99)	–14.29	–14.29
Pain	0.89 (.80–.97)	0.86 (.74–.99)	–25.0	–25.0
SI	0.81 (.66–.95)	0.82 (.62–1.0)	–25.0	–18.75
AOFAS				
Hallux MTP–IP	0.85 (.69–1.0)	0.80 (.54–1.0)	17.0	7.0
Ankle hindfoot	0.87 (.70–1.0)	0.81 (.50–1.0)	2.0	2.0
Midfoot	0.86 (.69–1.0)	0.77 (.46–1.0)	5.0	5.0
Lesser toe MTP–IP	0.83 (.67–.99)	0.81 (.55–1.0)	7.0	3.0
SF–36				
Physical function	0.75 (.58–.93)	0.96 (.93–1.0)	5.0	10.0
Role physical	0.65 (.48–.82)	0.72 (.48–.96)	25.0	25.0
Role mental	0.48 (.32–.64)	0.34 (.13–.54)		
Social function	0.60 (.41–.78)	0.60 (.32–.89)	12.5	12.5
Mental health	0.56 (.37–.74)	0.50 (.26–.74)	5.0	5.0
Energy/vitality	0.37 (.24–.51)	0.45 (.24–.66)		
Pain	0.63 (.49–.78)	0.68 (.46–.91)	0	11.1
General health	0.65 (.47–.83)	0.65 (.37–.93)	10.0	10.0

Area under the curve: value 1.0 = perfectly predictive, 0.5 = no better than chance (excessive noise), 0.0 = consistently incorrect. Values with confidence limits >0.5 have been highlighted.

differentiating between patients according to responses concerning pain, but CIs for the area under the curve were all much wider and, in the case of hindfoot and midfoot scales, crossed 0.5 in relation to foot-related problems. This indicated superior performance of the MOXFQ compared with the AOFAS in this regard. The SF-36 performed poorly in this analysis, with only the physical function domain achieving a reasonable area under the curve; the 95% CIs for the other domains crossed 0.5 and the sensitivity for all domains, including physical function, was low.

In terms of the pain transition item, best cut-points for the MOXFQ suggested that a change of 14 scale points for the walking/standing scale and 25 points for the pain and social interaction scales indicate a minimally important amount of change. Best cut-points for the AOFAS scales were 17 points for the hallux MTP-IP scale and 7 for the lesser toe MTP-IP scale; while for the SF-36 physical function domain the best cut-point was 5 scale points.

Comparison of the two anchor-based approaches

The best cut-point and MCID for each domain are shown in Table V. While the best cut-point is lower than the MCID only for the MOXFQ walking/standing scale, this is true of all AOFAS scales. This reflects the shallower ROC curves for the MOXFQ than for the AOFAS, with the MOXFQ thus providing poorer discrimination between patients reporting improvement vs no change. Nevertheless, the sensitivity and specificity of all domains in both instruments are high at between 70% and 80%. This contrasts with the domains of the generic SF-36 which all have low levels of sensitivity and which have best cut-points (where these could be calculated) greater than the MCID for most domains, with the exception of that for the Physical Function domain.

DISTRIBUTION-BASED MEASURES OF MINIMAL IMPORTANT CHANGE

The Cronbach's alpha, S.E.M., and MDC 90% for each scale within each instrument are shown in Table V. The Cronbach's alpha (representing internal reliability) was satisfactory (>0.7) or optimal (range 0.80–0.90)[35,36] for all MOXFQ and SF-36 domains, but for none of the AOFAS scales.

The error associated with a single use of the questionnaire, the S.E.M., is estimated as approximately 7, 8, and 12 scale points for the MOXFQ Walking/standing, Pain, and Social interaction scales, respectively. The MDC 90%, the value beyond which real change in the patient's health status is likely to have occurred, is estimated as 12, 15, and 20 for the three scales. These figures are comparable to the estimates for the SF-36 domains but are larger than those for the AOFAS scales.

COMPARISON OF THE MEASURES OF MIC FROM THE ANCHOR AND DISTRIBUTION-BASED APPROACHES

The results of the distribution (Cronbach's alpha, S.E.M., MDC) and anchor-based (MCID and ROC best cut-point together with its sensitivity and specificity) approaches to assessing MIC are shown in Table V for all the MOXFQ, AOFAS and SF-36 scales. The MCID:S.E.M. ratio is also shown.

The values for the SF-36 do not show good agreement between the two approaches, this is consistent with the generally low responsiveness and discrimination shown by this generic measure. Nevertheless, the MDC values are generally smaller than the MCID values for most domains within the two disease-specific measures, with the exception of the MOXFQ pain domain. This indicates consistency between the anchor and distribution-based approaches in the estimation of meaningful change since the MCID should

Table V

Cronbach's alpha scores and values (percentage scores) representing scaled S.E.M. and MDC for the MOXFQ, AOFAS and SF-36 scales. The MCID, in terms of those reporting their pain to be 'slightly better' at 12 months following foot surgery compared with before surgery is also shown, together with the MCID:S.E.M. ratio, and (differentiating between those who said their pain was 'Slightly better/much better/no pain now' vs 'no change/worse) the area under the ROC, the best cut-point, and values of sensitivity and specificity

	Distribution-based approaches			Anchor-based approaches					
	Cronbach's α	Scaled S.E.M	Scaled MDC 90%	MCID "slightly better"	MCID/ S.E.M.	ROC area (95% CI)	Best cut-point	Sensitivity	Specificity
MOXFQ									
WS	0.92	7.38	11.98	−16.07	2.18	0.82 (0.71−0.94)	−14.29	78.1	73.3
Pain	0.86	8.10	14.85	−11.82	1.46	0.89 (0.80−0.97)	−25.0	76.5	86.7
SI	0.73	11.92	20.14	−23.96	2.01	0.81 (0.66−0.95)	−25.0	73.1	80.0
AOFAS									
Hallux MTP−IP	0.63	9.26	7.10	24.75	2.67	0.85 (0.69−1.0)	17.0	75.0	87.0
Ankle hindfoot	0.57	7.30	6.30	8.90	1.22	0.87 (0.70−1.0)	2.0	90.9	92.8
Midfoot	0.59	8.34	6.74	12.40	1.49	0.86 (0.69−1.0)	5.0	81.8	88.2
Lesser toe MTP−IP	0.61	9.30	7.12	11.33	1.21	0.83 (0.67−0.99)	7.0	72.7	87.7
SF-36									
Physical function	0.88	7.74	14.52	12.78	1.65	0.75 (0.58−0.93)	5.0	64.3	82.1
Role physical	0.95	6.08	14.38	22.92	3.77	0.65 (0.48−0.82)	25.0	40.0	89.5
Role mental	0.94	5.86	16.30	4.86	0.83	0.48 (0.32−0.64)	–	–	–
Social function	0.86	8.70	24.33	7.29	0.84	0.60 (0.41−0.78)	12.5	38.5	81.6
Mental health	0.81	7.34	14.14	4.48	0.61	0.56 (0.37−0.74)	5.0	40.0	84.1
Energy/vitality	0.85	7.72	16.21	3.65	0.47	0.37 (0.24−0.51)	–	–	–
Pain	0.82	10.28	24.93	20.35	1.98	0.63 (0.49−0.78)	0	61.5	61.3
General health	0.80	8.43	15.15	9.58	1.14	0.65 (0.47−0.83)	10.0	53.9	87.9

ideally be greater than the MDC to enable its detection. The MCID estimates are, for the two disease-specific measures, on average around 1.8 times the S.E.M. (an average of 1.9 for the MOXFQ domains and 1.7 for the AOFAS domains).

Discussion

In this paper we have been able to demonstrate excellent responsiveness of both the MOXFQ and the AOFAS and, within these instruments, good agreement between the two approaches for estimating MIC. The generic SF-36 did not perform as well as the two disease-specific measures.

In terms of the practical uses of the measures calculated in this study, the MCID represents the mean change score of those patients who report feeling at least slightly better, whereas the best cut-point is the score above which patients are deemed to have at least slightly improved. The best cut-point may be used to determine, *post hoc*, whether an individual has, in terms of their response to the instrument, clinically significantly improved, while the MCID would be used, *a priori*, in calculating sample size. Any particular study should be large enough to have sufficient power (i.e., 80%) of detecting at certain alpha levels (e.g., $P < 0.05$) the MCID value, i.e., the smallest change patients consider to be meaningful. The results reported here are thus likely to be of benefit to those using and interpreting outcome measures in clinical practice, planning clinical trials and, in consultation with a statistician, in performing sample-size calculations.

RESPONSIVENESS

In terms of responsiveness and ES, it is important to note that the ES is not a property of a method of assessment but rather of the treatment intervention, and that it is influenced by the type of intervention, the amount of time from the intervention to the assessment and by the patient population[37]. Nevertheless, both of the foot-specific measures (MOXFQ and AOFAS) exhibited high levels of responsiveness that were superior to all domains of the generic SF-36. While the hallux scale of the AOFAS foot assessment appeared the most responsive overall, the summary measure of relative efficiency suggested that the MOXFQ was slightly more efficient than the AOFAS. Furthermore, for these two foot-specific measures, there were clear associations between domain scores and patients' global judgements of change in their foot pain and foot-related problems, while, with the exception of the physical function domain, this was generally not the case for the SF-36.

Neither of the radiographic measures of hallux valgus severity correlated with any measure of outcome, as also reported elsewhere[38] suggesting that radiographic measures are of limited use in the assessment of outcomes of hallux valgus surgery. A lack of correspondence between radiographic changes and peoples' symptoms has also been noted in other areas of orthopaedics, e.g., osteoarthritis of the hip and knee[39,40].

ANCHOR-BASED MEASURES OF MIC

The two specific measures were strongly, and linearly, related to the global transition items, enabling estimation of the MCID within those patients reporting that they felt 'slightly better'. These MCID estimates were equivalent to ES of 0.4, 0.2, and 0.8 (an average of 0.5) in the MOXFQ walking/standing, pain, and social interaction scales, respectively. In general, the ES were higher in the AOFAS, particularly in the highly specific hallux scale. Nevertheless, the ability of the different measures to differentiate between patients reporting different levels of improvement on transition, using ROC curves, indicated that the MOXFQ performed better than the AOFAS, while both of the foot-specific measures outperformed the SF-36 to a considerable degree.

COMPARISON OF ANCHOR AND DISTRIBUTION-BASED ESTIMATES OF MIC

The anchor and distribution-based approaches provided similar estimates of meaningful change, suggesting that for both foot-specific measures the MCID is around 1.8 times the S.E.M. This accords with estimates from a previous review of the relation between the MCID and S.E.M. in patients with chronic disease (a 1 S.E.M. correspondence) and back, lower neck and extremity pain (an approximately 2.3 correspondence)[41]. This concordance is important as the S.E.M. incorporates both the reliability and the variability of a measure. Agreement between the patient-reported MCID and the more mathematically derived S.E.M., and the MDC derived from it, gives weight to the relevance of the MCID in representing meaningful change.

It has previously been suggested that a one-S.E.M. criterion (or, equivalently, 0.5 SD when the reliability of the instrument is 0.75) can be used as a proxy for MIC in the interpretation of all health related quality of life instruments[42]. However, since the S.E.M. is simply a measure of detectable change it is more likely that this one-S.E.M. criterion is one for minimally detectable change than MIC[17]. The finding in this study that the MCID is, on average, 1.8 S.E.M.S is consistent with this and supports the validity of these MCID estimates.

MEASUREMENT PROPERTIES OF THE INSTRUMENTS

These findings provide strong support for the validity and responsiveness (and responsiveness is itself one indication of a measure's validity[43]) of both the disease-specific MOXFQ and AOFAS measures, suggesting that they perform well in patients undergoing hallux valgus surgery. All scales within these measures performed better (with greater precision) than the SF-36 in distinguishing between patients who gave different ratings of global change in their foot symptoms and satisfaction following surgery for hallux valgus.

However, while the generic SF-36 was much less efficient overall than either of the two disease-specific measures, the physical function and pain domains showed a statistically significant linear trend for greater degree of change with greater perceived improvement, and overall, the SF-36 physical domains performed better in patients receiving surgery for hallux valgus than the emotional domains.

Our evidence strongly suggests that a previous study, which also found little association between the SF-36 and the AOFAS, reached an erroneous conclusion in declaring the construct validity of the AOFAS to be poor[44], when it was the construct validity of the SF-36 *in the specific context of foot surgery* which was likely to be the more questionable in their study, the SF-36 not being designed to assess foot or ankle problems specifically, would therefore not be expected to perform with a high degree of responsiveness in the context of hallux valgus surgery and should not be used as the primary outcome measure in this context (as others have noted[45]). By contrast, the AOFAS clinical domains were designed specifically to assess foot or ankle problems and are very widely used for this purpose—despite the limited evidence, until now, for their reliability, validity or responsiveness[45].

Indeed, in addition to the stated aims of the paper, this study has provided evidence concerning the measurement properties of the four scales of the AOFAS clinical assessment. Here, the Cronbach's alpha (representing internal reliability) was less than satisfactory (<0.7) for all domains of the AOFAS[35,36]. However, the meaning of the Cronbach's alpha is difficult to interpret in relation to the AOFAS domains in view of the overlap and differences in scoring that are applied to some identical items contained in all four scales. Furthermore, the fact that some items are generic, being used in the calculation of all four scale scores, and others specific to that particular foot region may result in relatively lower levels of internal reliability. However, since power is a function of reliability and sample size, a less reliable instrument can achieve sufficient power if the sample size is increased[37].

CLINICAL IMPLICATIONS OF THE FINDINGS

Those wishing to investigate outcomes of foot and ankle surgical interventions have to choose which outcome measure(s) are the most appropriate to use. An important consideration is the measurement properties of the different measures being considered. We have here provided such details representing a few candidate measures, focusing mainly on the properties of the MOXFQ. However, other considerations are also of importance. Specific measures are less prone to interference—or 'noise'—from other conditions than are generic measures (e.g., the SF-36), although it is unrealistic to imagine that all noise can be eradicated[46]. The addition of a standard clinical assessment may add useful information to the overall assessment process, although the pros and cons of adding other measures need to be weighed carefully, as there tends to be a trade-off between the amount of data that is collected and the completeness of the data obtained.

In conclusion, the MOXFQ, devised with input from patients undergoing foot surgery, has previously been demonstrated to have very favourable measurement properties in terms of its reliability and validity[11]. By a number of different criteria, this measure has now also been shown to be highly responsive to clinical change in the context of hallux valgus surgery. This study provides scale properties which can be used to inform sample-size calculations for future trials. Further work is needed to determine the extent of its suitability as an outcome measure for other forms of foot and ankle surgery.

Acknowledgements

The authors would like to acknowledge the significant contribution, made to this study (including the contribution of ideas at every stage of the study and for performing all clinical assessments) by the following members of the Oxford and Birmingham Foot and Ankle Clinical Research Group: Grahame Lavis, B.Sc. (Hons), Consultant Podiatric Surgeon; Paul H. Cooke, F.R.C.S., Consultant Orthopaedic Surgeon, and Robert J. Sharp, F.R.C.S. (Tr & Orth) Orthopaedic Surgeon, Nuffield Orthopaedic Centre, Windmill Road, Oxford OX3 7LD & Mark L. Herron, F.R.C.S. (Tr & Orth) Consultant Orthopaedic Surgeon, The Royal Orthopaedic Hospital, The Woodlands, Bristol Road South, Birmingham B31 2AP.

We should like to thank The BUPA Foundation (the medical charity) for their generous support of this research project.

We should also like to thank all the patients who contributed their views and time to this study.

Appendix 1. The Manchester-Oxford foot questionnaire (MOXFQ)

Circle as appropriate: **RIGHT / LEFT FOOT**[1] ***During the past 4 weeks this has applied to me:***	*Please tick ✓ one box for each statement*				
	None of the time	**Rarely**	**Some of the time**	**Most of the time**	**All of the time**
1. I have pain in my foot	❑	❑	❑	❑	❑
2. I avoid walking long distances because of pain in my foot	❑	❑	❑	❑	❑
3. I change the way I walk due to pain in my foot	❑	❑	❑	❑	❑
4. I walk slowly because of pain in my foot	❑	❑	❑	❑	❑
5. I have to stop and rest my foot because of pain	❑	❑	❑	❑	❑
6. I avoid some hard or rough surfaces because of pain in my foot	❑	❑	❑	❑	❑
7. I avoid standing for a long time because of pain in my foot	❑	❑	❑	❑	❑
8. I catch the bus or use the car instead of walking, because of pain in my foot	❑	❑	❑	❑	❑
9. I feel self-conscious about my foot	❑	❑	❑	❑	❑
10. I feel self-conscious about the shoes I have to wear	❑	❑	❑	❑	❑
11. The pain in my foot is more painful in the evening	❑	❑	❑	❑	❑
12. I get shooting pains in my foot	❑	❑	❑	❑	❑
13. The pain in my foot prevents me from carrying out my work/everyday activities	❑	❑	❑	❑	❑
14. I am unable to do all my social or recreational activities because of pain in my foot	❑	❑	❑	❑	❑

15. During the past 4 weeks how would you describe the pain you usually have in your foot? *(please tick one box)*

None	Very mild	Mild	Moderate	Severe
❑	❑	❑	❑	❑

16. During the past 4 weeks have you been troubled by pain from your foot in bed at night? *(please tick one box)*

No nights	Only 1 or 2 nights	Some nights	Most nights	Every night
❑	❑	❑	❑	❑

[1] The foot to be assessed may be indicated here. Alternatively, each question may be customised to the right foot with all questions then repeated and customised to the left foot.

Finally, please check that you have answered every question

Thank you very much

References

1. Dawson J, Thorogood M, Marks S-A, Juszczak E, Dodd C, Lavis G, *et al*. The prevalence of foot problems in older women: a cause for concern. J Public Health Med 2002;24:77–84.
2. Balint GP, Korda J, Hangody L, Balint PV. Regional musculoskeletal conditions: foot and ankle disorders. Best Pract Res Clin Rheumatol 2003;17:87–111.
3. Wilder FV, Barrett JP, Farina EJ. The association of radiographic foot osteoarthritis and radiographic osteoarthritis at other sites. Osteoarthritis Cartilage 2005; 13:211–5.
4. Mann RA, Plattner PF. Ankle and foot: editorial overview. Curr Opin Orthop 1990;1:111–2.
5. Ferrari J, Higgins JP, Prior TD. Interventions for treating hallux valgus (abductovalgus) and bunions. Cochrane Database Syst Rev 2004;1:CD000964.
6. Parker J, Nester CJ, Long AF, Barrie J. The problem with measuring patient perceptions of outcome with existing outcome measures in foot and ankle surgery. Foot Ankle Int 2003;24:56–60.
7. Wright JG, Rudicel S, Feinstein AR. Ask patients what they want. Evaluation of individual complaints before total hip replacement. J Bone Joint Surg [Br] 1994; 76:229–34.
8. Schneider W, Knahr K. Surgery for hallux valgus. The expectations of patients and surgeons. Int Orthop 2001;25:382–5.
9. Pynsent P, Fairbank JTC, Carr A. Outcome Measures in Orthopaedics. Oxford: Butterworth-Heinemann 1993.
10. Royal College of Surgeons of England. Background to outcomes database. http://www.rcseng.ac.uk/surgical_research_units/ceu/projects/proj_outcomes.html. 2006.
11. Dawson J, Coffey J, Doll H, Lavis G, Cooke P, Herron M, *et al*. A patient-based questionnaire to assess outcomes of foot surgery: validation in the context of surgery for hallux valgus. *Published 'online first' (Sept)*. Qual Life Res 2006;15:1211–22, http://www.springerlink.com/content/1573-2649/?k=dawson; 2006.
12. Wilson DW. Hallux valgus and rigidus. In: Helal B, Wilson D, Eds. The Foot. London: Churchill Livingstone 1988:411–83.
13. Resch S, Stenstrom A, Jonsson K, Reynisson K. Results after chevron osteotomy and proximal osteotomy for hallux valgus: a prospective, randomised study. The Foot 1993;3:99–104.
14. Beaton DE. Understanding the relevance of measured change through studies of responsiveness. Spine 2000;25:3192–9.
15. Guyatt G, Walter S, Norman G. Measuring change over time: assessing the usefulness of evaluative instruments. J Chronic Dis 1987;40:171–8.
16. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. Control Clin Trials 1989;10: 407–15.
17. de Vet HC, Terwee CB, Ostelo RW, Beckerman H, Knol DL, Bouter LM. Minimal changes in health status questionnaires: distinction between minimally detectable change and minimally important change. Health Qual Life Outcomes;4, http://www.hqlo.com/content/4-1/54; 2006.
18. Wyrwich KW, Tierney WM, Wolinsky FD. Further evidence supporting an SEM-based criterion for identifying meaningful intra-individual changes in health-related quality of life. J Clin Epidemiol 1999; 52:861–73.
19. Fayers PM, Machin D. Quality of Life—Assessment, Analysis and Interpretation. Chichester: John Wiley & Sons Ltd, 2000.
20. Deyo RA, Centor RM. Assessing the responsiveness of functional scales to clinical change: an analogy to diagnostic test performance. J Chronic Dis 1986;39: 897–906.
21. Ware-JE J, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care 1992;30:473–83.
22. Jenkinson C, Stewart-Brown S, Petersen S, Paice C. Assessment of the SF-36 version 2 in the United Kingdom. J Epidemiol Community Health 1999;53: 46–50.
23. Kitaoka H, Alexander I, Adelaar R, Nunley J, Myerson M, Sanders M. Clinical rating systems for ankle-hindfoot, midfoot, hallux and lesser toes. Foot Ankle Int 1994;15:349–53.
24. Menz HB, Munteanu SE. Radiographic validation of the Manchester scale for the classification of hallux valgus deformity. Rheumatology 2005;44:1061–6.
25. SPSS statistical software. Release 13.0. Chicago, IL: SPSS Inc. (Headquarters) 2004.
26. StataCorp. Stata Statistical Software: Release 8.2. College Station, TX: Stat Corporation 2004.
27. Kazis LE, Anderson JJ, Meenan RF. Effect sizes for interpreting changes in health status. Med Care 1989; 27:S178–89.
28. Cohen J. Statistical Power Analysis for the Behavioural Sciences. New York: Academic Press 1997.
29. Liang MH. Evaluating measurement responsiveness. J Rheumatol 1995;22:1191–2.
30. Altman DG. Practical Statistics for Medical Research. London: Chapman & Hall 1991.
31. Stratford PW, Binkley JM, Riddle DL. Health status measures: strategies and analytic methods for assessing change scores. Phys Ther 1996;76:1109–23.
32. Kennedy DM, Stratford PW, Wessel J, Gollish JD, Penney D. Assessing stability and change of four performance measures: a longitudinal study evaluating outcome following total hip and knee arthroplasty. BMC Musculoskelet Disord;6, http://www.biomedcentral.com/1471-2474-6-3; 2005.
33. Hanley JA, McNeil BJ. A method of comparing the areas under receiver operating characteristic curves from the same cases. Radiology 1983;148:839–43.
34. Riddle DL, Stratford PW, Binkley JM. Sensitivity to change of the Roland-Morris Back Pain Questionnaire: part 2. Phys Ther 1998;78:1197–207.
35. Nunnally JC, Bernstein IH. Psychometric Theory. New York: Columbus OH: McGraw-Hill 1978.
36. Guyatt GH, Feeny DH, Patrick DL. Measuring health-related quality of life. Ann Intern Med 1993;118: 622–9.
37. Diehr P, Chen L, Patrick D, Feng Z, Yasui Y. Reliability, effect size and responsiveness of health status measures in the design of randomized and cluster-randomized trials. Contemp Clin Trials 2005;26:45–58.
38. Thordarson D, Ebramzadeh E, Moorthy M, Lee J, Rudicel S. Correlation of hallux valgus surgical outcome with AOFAS forefoot score and radiological parameters. Foot Ankle Int 2005;26:122–7.
39. Birrell F, Croft P, Cooper C, Hosie G, Macfarlane GJ, Silman A. Health impact of pain in the hip region with and without radiographic evidence of

osteoarthritis: a study of new attenders to primary care. PCR Hip Study Group. Ann Rheum Dis 2000; 59:857–63.
40. Dieppe PA, Cushnaghan J, Shepstone L. The Bristol 'OA500' study: progression of osteoarthritis (OA) over 3 years and the relationship between clinical and radiographic changes at the knee joint. Osteoarthritis Cartilage 1997;5:87–97.
41. Wyrwich KW. Minimal important difference thresholds and the standard error of measurement: is there a connection? J Biopharm Stat 2004;14:97–110.
42. Norman GR, Sloan JA, Wyrwich KW. Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. Med Care 2003;41:582–92.
43. Hays RD, Hadorn D. Responsiveness to change: an aspect of validity, not a separate dimension. Qual Life Res 1992;1:73–5.
44. SooHoo NF, Shuler M, Fleming LL. American Orthopaedic Foot and Ankle Society. Evaluation of the validity of the AOFAS clinical rating system by correlation to the SF-36. Foot Ankle Int 2003;24:50–5.
45. Button G, Pinney S. A meta-analysis of outcome rating scales in foot and ankle surgery: is there a valid, reliable, and responsive system? Foot Ankle Int 2004;25:521–5.
46. Dawson J, Fitzpatrick R, Murray D, Carr A. The problem of 'noise' in monitoring patient-assessed outcomes: generic, disease-specific and site-specific instruments for total hip replacement. J Health Serv Res Policy 1996;1:224–31.